DOCUMENT RESUME

ED 373 068 TM 021 789

AUTHOR Johnson, Randell G.
TITLE A Validity Generalization Study of the Multiple Assessment and Program Services Test.
PUB DATE Apr 93
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; *Grades (Scholastic); Meta Analysis; *Sample Size; Statistical Studies; *Test Validity
IDENTIFIERS Criterion Variables; Linear Models; *Multiple Assessment and Programs Services; Subtests; *Validity Generalization

ABSTRACT

The purpose of this paper was to research published validity reports of the Multiple Assessment and Program Services (MAPS) Test and to conduct quantitative analysis on the validity coefficients from the MAPS reports to determine the generalizability of the results and to identify which variables in the reports impact these coefficients. The specific question addressed was whether validity coefficients are related to the size of the sample populations, the subtest on the MAPS test, the location of testing sites, and criterion variable. This review used validity generalization procedures to evaluate the generalizability of previous test data from six studies (18 subtest cases). A general linear model was used to examine the relationships between the size of the populations, the subtest of MAPS, the location of testing sites, and the criterion variable used in determining the correlation coefficient. Analysis indicated that the validity coefficients were not generalizable across the different settings. Data from the individual states were significantly different from the national samples. Use of an alternate test as a criterion variable was significantly different from course grade and grade point average variables. Neither sample size nor the MAPS subtests were statistically significant. Eight tables are included. (Contains 26 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 068

A VALIDITY GENERALIZATION STUDY

OF THE

MULTIPLE ASSESSMENT AND PROGRAM SERVICES TEST

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY
RANDELL G. JOHNSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Randell G. Johnson, Ph.D.

Director of Evaluation and Planning
Gwinnett Technical Institute
1250 Atkinson Rd./P.O. Box 1505
Lawrenceville, Georgia 30246-1505

Paper presented to the American Educational Research Association
Distinguished Paper Session
April 14, 1993
Atlanta, Georgia

TM021789

BEST COPY AVAILABLE

Table of Contents

Page

Abstract ii

Introduction 1

Statement of the Problem 4

The Theory of Validity Generalization 6

Purpose 7

Limitations 7

Review of the Literature 9

History of the Development and Use of MAPS 9

MAPS Validity Studies 11

Validity Generalization 12

The Concept of Validity Generalization 13

Methodology 14

Results 15

Analysis of the Data 15

Validity Generalization Results 15

Results for General Linear Model 16

Conclusion 18

Discussion 21

References 23

Tables 27

ABSTRACT

The purpose of this paper was to research published validity reports of the Multiple Assessment and Program Services (MAPS) Test and to conduct quantitative analysis on the validity coefficients from the MAPS reports to determine the generalizability of the results and to identify which variables in the reports impact these coefficients.

Previous research on the MAPS test includes many different school districts and populations in several states, and validity studies have been conducted on the MAPS data gathered within these different school populations. This study addressed the following research question: Are the validity coefficients related to the size of the sample populations, subtest on the MAPS test, the location of the testing sites, and criterion variable? This review used validity generalization procedures to evaluate the generalizability of previous test data. A general linear model was used to examine the relationships between the size of the sample populations, the subtest of MAPS, the location of the testing sites, and the criterion variable used in determining the correlation coefficient.

A validity generalization analysis indicated that the validity coefficients were not generalizable across different settings. The data collected in the individual states were significantly different from the national samples. Also, the use of an alternate test as a criterion variable was significantly different from course grade and grade point average variables. Neither the sample size nor the MAPS subtests were statistically significant.

A Validity Generalization Study of the Multiple Assessment and Program Services Test

Standardized tests play a major role in the admission processes of most colleges and universities. In 1982, 86% of four-year public colleges and 96% of 4-year private colleges considered "standardized test scores an important factor in admission" (Amberg, 1982, p. 536). The interpretation of these scores becomes very important because institutions rely on the results of these tests to assist them in placing their applicants in college courses, developmental studies, or remedial work. The institutions spend a great deal of time and money in validating the cut-off scores for accepting and placing applicants.

During the 1980s, researchers who made these inferences from criterion-related test scores "were encouraged to conduct . . . local validity studies" because of the general belief that the "validity of an inference from a test . . . should be situation specific" (Mehrens and Lehmann, 1987, p. 98). Situation-specific testing has been the standard. Noeth (1976) said that "one of the most efficient uses of test data is for the local schools to conduct their own validity studies" (pp. 60-61). This view was based on the belief that populations and, thus, the validity coefficients would vary greatly from one situation to another (Mehrens and Lehmann, 1987).

Institutions, which use criterion-referenced tests as part of their admission process, collect data for validity studies on a periodic basis to confirm their use of the scores for placement. Deciding (or being mandated) to change standardized tests, the institutions must conduct validity testing, set new cut-off scores, and review/revise the admission

process. Thus, the validation of new criterion-referenced tests require an enormous investment of time and money for an institution. Often, changing tests requires institutions (1) to conduct testing and complete test analysis before the test is officially adopted or (2) to open the admission/placement policy until data are collected for analysis.

In 4-year colleges and universities, validity testing is usually conducted on sample populations before the test is incorporated into the admission process. Technical and community colleges are faced with a more critical dilemma. Since most of these postsecondary institutions maintain an "open door" admission policy, the test's predictive ability in accurately placing students into courses and programs of study is of paramount importance. It is, however, these institutions which are more often required to change tests or to modify their admission practices and are often less able to afford the preliminary testing.

For example, in 1989, Georgia's Department of Technical and Adult Education (DTAE), the governing board for postsecondary technical institutions, investigated a new criterion-referenced test for all of the technical institutes in the state. Formerly, no single instrument was required for placement in the more than thirty institutes. Without conducting validity testing, DTAE recommended four subtests (Reading, Language, Cumulative, and Elementary Algebra) of the College Board's Multiple Assessment and Program Services (MAPS) tests for use state wide as the placement test of choice. No data were collected before Spring 1990 when DTAE required that these tests be used in all technical institutes in the state. DTAE did, however, make arrangements with the

College Board to collect data as the test was administered and to analyze the data periodically.

Another example occurred in Florida in 1984. Florida legislation (Postsecondary Education Act, FL. Code Ann. 240. 117-118, 1985) mandated that by June 30, 1984 the State Board of Education would specify common placement tests for use by university and community colleges in assessing basic communication and computation skills of all students planning to enter those schools. Additionally, the state required cut-off scores on these MAPS tests by July 1, 1985 for determining if students needed extra preparation to acquire basic college skills.

Georgia and Florida are only two examples of problems related to localized test validity faced by postsecondary institutions which require placement testing. Arizona's Maricopa County Community College District (MCCC) is one system which took the time to collect the MAPS data prior to test selection and then rejected the use of MAPS as a college placement test (Abbott, 1986). In 1985 and 1986, MCCC conducted a project to develop a district-wide database for use in decisions about policies, programs, and procedures as related to student assessment, advisement, and placement. The report states that the purpose of the assessment was to place students in courses which the assessment instrument indicated a high likelihood of their academic success. Predictive validity coefficients were low to moderate for the MAPS test. MCCC District confirmed, over a period of two years, that their data were inconclusive to decide on a placement test; therefore, they made a selection based on the coordinator's opinion survey.

In each of these case, the states did not consider accepting the results of nation-wide validity studies conducted by the College Board to confirm the MAPS test as a valid placement test in a local setting. These states may be suggesting that the local states are very different from one another and, thus, different from the national samples as a whole. They also may be conducting, as did Arizona, a comparison of placement tests.

Statement of the Problem

Educational systems and institutions collect data to validate their use of test scores in an admission/placement process. Since 1979, 57 publications have been written on the development or use of the MAPS tests. Fifty-one of these publications are reports of validity testing. The College Board published 18 of them to describe the tests' development and validity testing or services by the College Board which support the use and continued validity testing of MAPS; however, the College Board was unable to supply the raw data or the summary data to support the conclusions reported in these publications. A total of five studies reported the validity coefficients which included the MAPS tests of Reading, Language, Computation, and Elementary Algebra. After the College Board placed the MAPS tests and its other placement tests under the "MAPS Umbrella" (The College Board, 1980), it began the Assessment and Placement Services for Community Colleges which conducts local validity studies on these tests at the expense of the local institution, and the results are kept confidential and unpublished (The College Board, 1986). In other words, the College Board is charging individual systems and institutions for conducting localized validity testing and, under the veil of confidentiality, are not sharing the tests' validity with the general consumer.

Thirty-three of the publications were published reports by individual states or institutions presenting data to support the tests' use, continued use, or the validity of cut-off scores; however, only Arizona (Abbott, 1986) published the validity coefficients. Additionally, the test is used at many similar institutions which periodically collect their own validity data but do not publish the results or include their validity coefficients in their publications.

If Georgia's technical institutes could have used Florida's community college validity data on MAPS, Georgia would have saved a great deal of time, effort, and money which is being used in conducting validity studies on data collected state-wide. In order to determine whether or not validity data can be used in different settings, there is a need for an empirical process to confirm that validity data are generalizable from one situation to another and from one state to another. This would eliminate what sometimes becomes a long waiting period before a placement test can be instituted as well as the cost of collecting the locally conducted validity studies (Mehrens and Lehmann, 1987). This generalizability would also allow states to begin using placement tests state-wide before institutions conduct local validity studies.

If the data from earlier validity studies transfered to other settings, then preliminary validity studies could be eliminated and ongoing data could be collected and analyzed to help fine tune the placement practices in order to best serve the students at local institutions.

The Theory of Validity Generalization

Validity generalization (VG) was introduced in the mid-1970's and provides a systematic framework for examining the "degree to which inferences from scores on tests, can be transported across different situations" (Burke, 1984, p. 94). VG theory provides a framework for examining properly conducted validity studies and the extent to which results can be generalized across institutional lines and state lines. According to Schmidt (1985), one of the developers of this theory, validity generalization focuses on estimating the true variance of study correlations and effect sizes. At that time, VG procedures had been applied in the analysis of over 500 research areas related to employment selection, and each one represented a predictor-job performance combination (Schmidt, 1985).

Although the belief that the validity of inferences from test scores should be situation-specific is beginning to change, institutions making such inferences are still encouraged to conduct their own local validity studies. While the theory of validity generalization does not substantiate some current beliefs that "local validation is no longer necessarily required," it does "support a claim of validity in a new situation" (Mehrens and Lehmann, 1987, pp. 98-99).

Until sufficient data are collected for institutions to validate the MAPS test for their own specific populations, the theory of Validity Generalization suggests that studies conducted in other states can help to provide evidence for immediate use regarding the validity of the test scores. Through VG, researchers can examine validity studies conducted across the country and identify the extent to which these inferences are generalizable to educational situations in other states.

Purpose

With such a high dependence on placement testing for first-time-in-college students, institutions reluctantly spend the time and money to conduct local validity studies or hire the College Board, the developer and marketing enterprise for the tests, to conduct local validity studies for them. Additionally, it also appears that validity studies, which the College Board conducted nationally, and the results of studies conducted in other states should be useable in local validity studies.

The purpose of this study was to examine previous validity studies which had been conducted on the MAPS test and to examine whether or not their inferences could be generalizable. Additionally, this study examined the components of these studies to determine the relationship(s) between the validity coefficients and the different variables in the studies.

In an attempt to fulfill this purpose and to respond to the questions left unanswered by the literature, this report was guided by a sequence of research questions:

1. Are the validity coefficients related to the size of the sample populations?
2. Are the validity coefficients related to the type of subtest?
3. Are the validity coefficients related to the location of the testing site?
4. Are the validity coefficients related to the criterion variable?

Limitations

This study investigated the validity tests conducted on MAPS which were available in the literature and in research reports from Arizona, Florida, New Jersey, and Tennessee, which use the test state wide.

Seven different studies (containing 21 cases) had been conducted nation wide which provided the validity coefficients necessary to conduct a validity generalization; however, one study (which provided validity coefficients for three subtest cases and criterion variables) lacked sufficient sample size information. Therefore, this study included only the six studies which provided sample sizes and validity coefficients for 18 subtest cases.

Five of the six studies were conducted by the College Board between 1975 and 1985, reporting their results in the technical manuals for MAPS; however, the College Board could not make the original data or summary data from these studies available for use in this study. The sixth study was conducted in Arizona.

Review of the Literature

History of the Development and Use of MAPS

Tests in Print (Mitchell, 1983), traditionally used by researchers and students seeking information regarding tests, offers the following brief description of the Multiple Assessment Programs and Services (MAPS) tests:

> MAPS was designed to help colleges make decisions about placement levels and remediation needs of entering as well as continuing students. MAPS provides data in the assessment areas of remediation, placement, exemption, selection, instruction, guidance, and counseling. MAPS is composed of three biographical questionnaires and 60 tests which were derived from programs already in use. The programs listed are "Comparative Guidance and Placement Program, Descriptive Tests of Mathematics Skills, Instructional Admissions Testing Program, Institutional Test of Standard Written English, and Testing Academic Achievement." The MAPS program is administered by the College Board and Educational Testing Service (p. 266).

The College Board maintains that MAPS scores should be used for placement and remediation needs rather than for making decisions on whether to admit students to college (Mitchell, 1983). The College Board (1987a), indicates that the purpose of MAPS is to make available to colleges tests for assessing the needs of students entering college for the first time. Test scores are used for placement of students into appropriate levels of remedial, developmental, regular, or advanced courses.

The College Board (1980) consolidated under one MAPS "umbrella" a wide variety of tests and questionnaires that are useful in the placement process. Included are the Scholastic Aptitude Test (SAT) and tests of ability to do academic work on an introductory college level such as English Composition, Mathematics Level 1, American History and Social Studies, Biology, Physics, French Reading, German Reading, and Spanish Reading. Although these tests may be used independently and may contain subtests of their own, they are offered under the College Board's "umbrella" so that institutions can select the most appropriate test for them. Included under that same umbrella is a set of descriptive tests of skills in reading, writing, computation, elementary algebra, and intermediate algebra which are designed for level I institutions such as community colleges and technical institutes. This paper focused on this set of descriptive skills tests which has been adopted by many level I institutions in several states and which is commonly referred to as the MAPS test.

The MAPS Placement Research Service was presented in 1986 as a new service designed to help colleges use the different MAPS tests. Test scores and criterion data supplied by the colleges are analyzed by the service and reports are sent to the colleges (Livingston, 1986). Most colleges are interested in such analyses as the relationship between students' preadmission test scores and their college course grades (The College Board, 1986). These types of analyses help institutions (which supply data to the service) to discover which tests accurately predict success in certain courses and which score levels should be used for placement of students at different skill levels. Reports to schools include score distributions of predictor and criterion measures, two-way tables of score

intervals on predictor and criterion variables, and tables to predict a student's criterion score from two or more predictor scores. The service uses a step-wise procedure to determine which variables contribute the most variance to the correlation between predicted and actual criterion scores. This information is confidential and for the individual school's use; therefore, the College Board does not make available validity correlations to institutions considering the adoption of the tests.

The data for this study were collected from the College Board's national testing of the original forms of the skills tests and from those three states which agreed to share the results of their validity studies.

MAPS Validity Studies

MAPS reveals considerable change over its short history. The College Board has initiated new tests and services which adapt to the needs of a growing and changing population of students. Most of the validity studies which are applicable to current users of the MAPS test have been reported in the last 8 to 10 years. These studies have correlated MAPS scores (Form A) to such criterion variables as its alternate test (Form B), course grades, or grade point averages.

In the last ten years, the College Board and other educational systems across the United States have conducted validity studies. These data from these studies represent a cross-section of first-time-in-college students. Some states have conducted their own situation-specific tests. Four of these states are Arizona, Florida, New Jersey, and Tennessee. Three of the four states currently use the test; Arizona chose the ASSET test.

By their actions, these states support the practice of doing local studies to establish norms and cut-off scores for their particular institutions. Reports of such studies came from Miami-Dade Community College (M-DCC) in Florida (Davis, 1985), the Tennessee State Board of Regents (SBR) (Nicks, 1985), and the New Jersey Basic Skills Council (1988). The College Board (1986) Assessment and Placement Services for Community Colleges supplies national norms and consultants to assist colleges in establishing local norms and cut-off scores. The lack of published reports indicates that individual colleges are either applying the national norms and deriving cut-off scores for local application rather than conducting local studies or are not publishing their validity data.

Validity Generalization

Historically, institutions and systems making criterion-related validity inferences from the test scores were encouraged to conduct their own local validity studies because validity inferences from a test should be situation specific (Mehrens and Lehmann, 1987). The reasons for this dominant belief are that the correlations often varied across settings, and the correlations were often low. Many analysts have challenged the need for situation-specific studies. These analysts attribute the variation in correlations to statistical artifacts, such as small sampling error, criteria and test unreliability, and restrictions in the test score ranges (Hirsh, Northrop, & Schmidt, 1986; Mehrens and Lehmann, 1987; Linn, Harnisch, & Dunbar, 1981; Schmidt, 1985). "It need not be concluded, however, that all of the variability between studies in validities is attributable to statistical artifacts for the idea of validity generalization to be useful" (Linn, et al.,

1981, p. 282). Discovering in the study that as much as 70% of the variance in observed validity is attributable to statistical artifacts, the researchers conclude that the generalizability of validity is more than adequate to "support the conclusion that the true validity is nonzero without the need for a situation-specific study" (Linn, et al., 1981, p. 288).

The Concept of Validity Generalization

Validity generalization is the "degree to which inferences from scores on tests, can be transported across different situations" (Burke, 1984, p. 94). This concept is a practical part of situation-specific validity studies where it is practical to assume that the predictive validity for one class is applicable to the next class. That is, admission criteria for a new class are derived from the results of the prior class because criterion data are unavailable for the new class. This is a type of widely accepted generalizability (Linn, et al., 1981).

In academic settings, frequently used admissions tests are correlated with first-year grade averages or specific course grades and cut-off scores are set for acceptance. This practice is based on the concept that there is no significant difference between the validity coefficients over the different populations although variations in the observed correlations occur from class to class and from school to school. In very large validity studies, these results are generalized for an entire state or from state to state. If an analysis of the validity coefficients from different populations were conducted and no significant differences identified, then it suggests that the validity coefficient can be generalized across different populations.

Variations do exist in the observed correlations; however, a number of statistical artifacts are believed to "influence the size and variability of observed validity coefficients" (Burke, 1984, p. 95). Validity generalization can be viewed as the application of meta-analysis to the problem of examining validity evidence across settings (Hedges, Shymansky, & Woodworth, 1989). The focus is on controlling statistical artifacts and estimating the variance of the effect size. Examples of statistical artifacts include predictor reliability, criterion reliability, range restriction in the predictor, and sample size (Burke, 1984).

Methodology

Validity studies conducted across the United States between 1975 and 1988 provided this study with all of the available correlation coefficients between the identified MAPS subtest scores and a criterion variable. This data is presented in Table 1. To examine whether or not the inferences from these studies could be transferred to other states, a validity generalization study was conducted to analyze these validity coefficients.

The z statistic was used to normalize the distribution of the validity coefficient, r, and to make variance independent of the population correlation (ρ) (Hedges & Olkin, 1985). A general linear model was used to examine the relationship between the validity coefficients and sample size, subtest on the MAPS, location of the testing sites, and criterion variable used in determining the coefficient.

Insert Table 1 About Here

Results

Analysis of the Data

The 18 validity coefficients used in this study were the result of six studies conducted over a period of thirteen years. Studies one through four were conducted by the College Board between 1975 and 1988 using national samples (The College Board, 1985; 1986; 1988a; 1988b). Study five was conducted in New Jersey by the College Board (The College Board, 1987b). Arizona's Maricopa County Community College District conducted study six and published its results in 1986 (Abbott, 1986).

In this study, the indicated sample size was the number of students (or approximate number) given in the study. Some of the larger studies rounded sample sizes to the nearest hundred or estimated the sample size. Also, the validity coefficient was the correlation between students' scores on the MAPS subtest and the various criterion variables such as course grade, grade point average, or scores on an alternate form of the MAPS test.

Validity Generalization

There is no one specific process or formula by which validity generalization results can be achieved and applied; however, it is important that the process control for within-study variability which can account for up to 70% of the variance of any study (Linn, et al., 1981). Hedges and Olkin (1985) developed a method of combining estimates of correlation coefficients in studies where the sample sizes are large and calculating approximations to the distribution of the sample correlation coefficient. They use the z

transformation in order to normalize the distribution of correlation r and to make the variance independent of mean p. The formula used in this transformation was:

$$z = z(r) = \frac{1}{2} \log \frac{1 + r}{1 - r}$$

The validity coefficients in this study were transformed before analysis. Table 2 presents the means and standard deviations by study characteristics.

Insert Table 2 Here

Results for General Linear Model

This study used a general linear model to investigate the relationship between the validity coefficients and the size of the different sample populations, subtest of MAPS, location of the testing site, and criterion variable.

The Pearson Product-Moment Correlation of the z coefficients with sample size was not significant, r (16) = .029, NS. Using the table of critical values for the Pearson Product-Moment Correlation Coefficient (Shavelson, 1981), the critical value for a sample of 18 is .4438 at the .05 level of significance.

Using each remaining independent variable in a separate linear model with the z coefficient as the dependent variable, the analysis of variance (ANOVA) produced significant results for location and criterion variable and not for subtest. Tables 3, 4, and 5 present the summary data for the analysis of variance by subtests, location, and criterion variables on the z coefficients calculated using SYSTAT (Wilkinson, 1986).

Insert Tables 3, 4, & 5 About Here

Since the F-values were significant for two independent variables, location and criterion variable, Tukey HSD tests were conducted for all possible pairwise comparisons between the means with an overall level of significance alpha = .05. Tukey's HSD tests identify where the differences occurred which gave rise to the significant F-value (Shavelson, 1981).

Using the location variable (Table 6), the Tukey HSD comparison of means of Arizona and New Jersey indicates that they are not significantly different at the .05 alpha level. These two states used basically the same subtests and an alternate test scores as their criterion variables. The means of the national sample are significantly different from the individual states at the .05 alpha level. The state-level validity coefficients are larger than the national mean, and the standard deviation was larger for the national sample. The sample sizes were much larger than the individual state samples; however, different combinations of subtests were administered in the determination of different criterion variables.

Insert Table 6 About Here

Using the criterion variable (Table 7), the Tukey HSD comparison of means for course grade and grade point average indicates that they are not significantly different at the .05 alpha level. The means of the alternative test were significantly different from

course grades and grade point averages at the .05 alpha level. The purpose of alternative tests are to measure the same criteria as the subtests in the study.

Insert Table 7 About Here

Conclusion

Reports from Arizona, Florida, Tennessee, and New Jersey agree that MAPS discriminates between students who need remedial or developmental (R/D) courses and those who possess the skills necessary for regular college courses (Abbott, 1986; The College Board, 1987b; Davis, 1985; Davis, Kaiser, & Bone, 1987; Mitchell, 1983). Most testing reports from the various states agree that MAPS is an effective placement test and should be used for that purpose.

Although they use different cut-off scores, the states agree that the MAPS tests effectively identify those students who need R/D courses. They also have a common concern for the high percentage of students who score below cut-off on each test. The New Jersey Basic Skills Council (1989) said (regarding fall 1989 test results) that the size of the 'Lack Proficiency' category continues to concern a higher education system which is striving toward excellence. The fact that other states with large testing programs typically report similar or lower results is of little consolation. Davis (1985) reported that two-thirds of fall 1985 M-DCC freshmen tested into some form of college preparatory work. In Tennessee, Davis, Kaiser, and Boone (1987) expressed concern that large numbers of under-prepared students are applying to SBR institutions.

The MAPS tests assists level I colleges to place their applicants in remedial, developmental, or college programs of study according to their predicted ability to be successful at those levels. However, many institution, systems, and state boards of education do not use the tests because of the enormous costs for localized testing that is required to establish test validity for the local population. In some cases, the money and the time are not available to obtain these preliminary test results.

Many of these institutions use the services of the College Board and use their reported data to establish minimum cut-off scores for placement into specific programs of study. Dr. Lucky Abernathy, Director of the MAPS Program, says that the College Board's services with an individual institution can be of nominal cost; however, since the College Board is compensated for conducting these studies, the results are placed into confidential files. The individual institutions, however, are free to share the information and/or publish it. For those institutions which have enough professional staff to conduct local validity studies, they are seldom published and/or shared with other institutions or systems because the results are often considered situation specific and of no use to other populations. Because of these circumstances, this study was limited to only 18 cases with validity coefficients and enough data to help determine if the findings of these studies are generalizable across populations. It was important that this study employ empirical processes to address the generalization of the test results across different states, to identify artifacts or variables which invalidate its use across populations for college placement, and to examine the relationship between the correlation coefficients and the independent variables in each study.

To test the generalizability of the results of the different studies across different populations, validity generalization was used. In this meta-analytical approach to generalizability, the z transformation of the validity coefficients and the asymptotic n were used to reduce the impact of statistical artifacts even though the sample populations for the 18 cases were moderate to large. This empirical process suggested that these test results could not be transported across different situations.

Since many analysts attribute the variations in correlations to statistical artifacts (Mehrens & Lehmann, 1987), the relationships between the validity coefficients and the independent variables (sample size, MAPS subtest, testing location, and criterion variables) were examined by using a general linear model.

To determine how the validity coefficients related to the sample size, a Pearson Product-Moment Correlation Coefficient was used. The Pearson correlation of the sample size to the z coefficient indicated no significant correlation; however, the significant F-value of the ANOVA on location indicates that the difference in population means from state to state was probably due to treatment and did not arise from sampling error. The different sample sizes and the different criterion variables used are examples of the problems with the data reported from the College Board.

A general linear model investigated the relationship between the validity coefficients and the subtest of MAPS, location of the testing site, and criterion variable. Although there is no significant difference in mean validity coefficients related to the sample size, the location of the testing site and the criterion variable are significantly different. A Tukey HSD delineated exactly where the difference occurred for each

independent variable. The studies representing national testing locations varied significantly from the states of Arizona and New Jersey. Also, of the three criterion used in the different studies, those studies which used the alternative test as their criterion variable were significantly different from those using course grades and grade point averages.

Discussion

The results of this review of the literature on MAPS confirms the earlier research conducted in Arizona, Florida, New Jersey, and Tennessee. The MAPS tests in reading, writing, computation, and algebra skills which are used in level I technical and community colleges are effective tests for placing first-time-in-college students into remedial, developmental, and college programs. The strength of these conclusions is, however, based on only a handful of reports from limited systems because the majority of the studies are conducted for individual systems by the College Board and the individual institutions do not publish or share the results. It was only in the cases of the few institutions or systems which published information, and of course the College Board publications, that data were available.

Because of the limited amount of information available, it appeared prudent to use that information fully which included a quantitative analysis. A validity generalization analysis indicated that the validity coefficients from the 18 studies were not generalizable (i.e., transferrable across different situations). Tukey HSD tests identified where differences occurred in the general linear models which were conducted on all independent variables. The national sample data (the College Board technical manual's

basis for the test's validity) are significantly different from New Jersey and Arizona studies which were not significantly different. The College Board could not make its data or summary data available for this study.

In summary, the results of this study should be viewed as more suggestive than definitive. Although the 18 studies have large sample populations of first-time-in-college students, the sample sizes vary greatly from study to study. Additionally, the MAPS subtests were all administered and considered in the determination of the validity coefficients. However, fewer calculation/algebra tests were considered than reading and writing tests. These studies should be replicated using relatively equal sample sizes, testing the subtests equally, and using the same criterion variable.

One of the most efficient uses of this study can be to encourage the College Board (and other agencies which verify nationally normed instruments which are used as placement tests for first-time-in-college students) to conduct nation-wide tests with equal sample sizes, to use the same subtests and criterion variables, and to maintain the raw or summary data so that it can be shared with state and local institutions and systems. Also, local institutions, systems, and states should consider conducting situation-specific validity studies so that the data can be shared or published for the purpose of research studies similar to this one.

REFERENCES

Abbott, J. A. (1986). Student assessment pilot project: Maricopa County Community College District (JCEP Project #JZ-309, 1985-86). Phoenix, AR: Maricopa County Community College District. (ERIC Document Reproduction Service No. ED 170 154)

Amberg, J. (1982). The SAT. The American Scholar, 51, Autumn, 535-542.

Burke, M. J. (1984). Validity generalization: A review and critique of the correlation model. Personnel-Psychology, (Spring, 1984), 93-115.

The College Board. (1980). How MAPS can help you with the admissions selection process. New York: Author.

The College Board. (1985). Assessment and placement services for community colleges: Using and interpreting scores. New York: Author.

The College Board. (1986). MAPS placement research service. New York: Author.

The College Board. (1987a). How MAPS can help you with placement. New York: Author.

The College Board. (1987b). The New Jersey college basic skills placement test program: Your information base for outcomes assessment. New York: Author.

The College Board (1988a). User's guide to the descriptive tests of language skills. New York: Author.

The College Board (1988b). User's guide to the descriptive tests of mathematics skills. New York: Author.

Davis, D. (1985). MAPS entry-level Basic Skills Testing outcomes for first-time-in-college students at Miami-Dade Community College fall term 1985 (Report No. 85-41). Miami, FL: Miami-Dade Community College.

Davis, T., Kaiser, R., & Boone, J. (1987). Speededness of the Academic Assessment Placement Program (AAPP) Reading Comprehension Test (Report No. 2). Memphis, TN: Memphis State University.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). Modern methods of meta-Analysis. Washington D.C.: National Science Teachers Association.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-Analysis. Orlando: Academic Press, Inc., 1985.

Hirsh, H. R., Northrop, L. C., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. Personnel Psychology, 39, 399-420.

Linn, R. L., Harnisch, D. L, and Dunbar, S B. (1981). Validity generalization and situational specificity: An analysis of the prediction of first-year grades in law school. Applied Psychological Measurement, 5(3), 281-289.

Livingston, S. A. (1986). The MAPS Placement Research Service. New York: The College Board. (ERIC Document Reproduction Service No. ED 272 518)

Mehrens, W. A., Lehmann, I. J. (1987). Using standardized tests in education. (4th ed.). New York: Longman.

Mitchell, J. V., Jr. (Ed.) (1983). Tests in print III. Lincoln, NE: University of Nebraska, Buros Institute of Mental Measurement.

New Jersey Basic Skills Council. (1988). Report to the Board of Higher Education on the results of the New Jersey College Basic Skills Pacement Testing Fall 1987 entering freshmen. New Jersey: State Board of Higher Education.

Nicks, R. S. (1985). Approved placement scores. Nashville, TN: Tennessee State Board of Regents.

Noeth, R. M. (1976). Converting student data to counseling information. Measurement and Evaluation in Guidance, 9 (2), 60-69.

Postsecondary Education Act, FL. Code Ann. 240.117-118 (1985).

Schmidt, F. L. (1985). From validity generalization to meta-Analysis: The development and application of a new research integration procedure. Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985). (Eric Document Reproduction Service No. ED 255 554)

Shavelson, R. J. (1981). Statistical Reasoning for the Behavioral Sciences. Boston: Allyn and Bacon, Inc.

Wilkinson, L. (1986). SYSTAT: the system for statistics. Evanston, IL: SYSTAT, Inc.

Table 1

Validity Studies of Multiple Assessment Program and Services

Study	Sample Size	Validity Coefficients	z	MAPS Subtest	State	Criterion Variable
The College Board (1985)	640	.28	.125	1	3	2
	6400	.32	.144	2	3	2
	2900	.43	.200	3	3	2
The College Board (1986)	307	.29	.130	1	3	1
	306	.20	.088	2	3	1
	257	.28	.125	3	3	1
The College Board (1988a)	1100	.88	.597	1	3	3
	571	.88	.597	2	3	3
The College Board (1988b)	803	.84	.530	3	3	3
	467	.80	.477	4	3	3
	297	.81	.489	5	3	3
The College Board (1987)	6000	.89	.618	1	2	3
	6000	.91	.663	3	2	3
	6000	.92	.690	4	2	3
Abbott, J.A. (1986)	1939	.90	.639	1	1	3
	1046	.87	.579	3	1	3
	1046	.91	.663	4	1	3
	1046	.86	.562	5	1	3

NOTE.

The **MAPS Subtests** were coded as follows:
1 = Reading, 2 = Writing, 3 = Computation,
4 = Elementary Algebra, and 5 = Intermediate Algebra

The **States** were coded as follows:
1 = Arizona, 2 = New Jersey, 3 = National Samples

The **Criterion Variables** were coded as follows:
1 = Course Grade, 2 = Grade Point Average,
3 = Alternate Test Score

Table 2

Means and Standard Deviations by Study Characteristics

	N	Validity Coefficient Mean	Validity Coefficient SD	z Mean	z SD
Subtest					
Reading	5	.648	.331	.422	.269
Writing	3	.467	.363	.276	.279
Computation	5	.666	.290	.419	.241
Elem. Algebra	3	.877	.067	.610	.116
Inter. Algebra	2	.835	.035	.526	.052
Location					
National Sample	11	.546	.289	.318	.215
New Jersey	3	.907	.015	.657	.036
Arizona	4	.885	.024	.611	.048
Criterion Variable					
Course Grade	3	.257	.049	.114	.023
Grade Point Average	3	.343	.078	.156	.039
Alternate Test Score	12	.873	.039	.592	.068

Note. Z statistic was used to normalize distribution of r and to make variance independent of the population correlation. The formula used in this transformation was:

$$z = z(r) = \frac{1}{2} \log \frac{1 + r}{1 - r}$$

Table 3

Summary of ANOVA for MAPS Subtests using Transformed (z) Scores

Source of Variation	Sum of Squares	Mean Square	DF	F-Value
Subtest	.185	.046	4	.853
Error	.707	.054	13	
TOTAL	.892	.100	17	

*$p < .05$

Note. Critical value for F is 3.18 (alpha = .05)

Table 4

Summary of ANOVA for Testing Locations by State using Transformed (z) Scores

Source of Variation	Sum of Squares	Mean Square	DF	F-Value
State	.421	.210	2	6.689*
Error	.472	.031	15	
TOTAL	.893	.241	17	

*$p < .05$

Note. Critical value for F is 3.63 (alpha = .05)

Table 5

Summary of ANOVA for Criterion Variables using Transformed (z) Scores

Source of Variation	Sum of Squares	Mean Square	DF	F-Value
Criterion	.837	.418	2	113.168*
Error	.055	.004	15	
TOTAL	.892	.422	17	

*$p < .05$

Note. Critical value for F is 3.68 (alpha = .05)

Table 6

Tukey HSD Comparison of Location Means

	National X = .318	Arizona X = .611	New Jersey X = .657
National X = .318	----	.293*	.339*
Arizona X = .611		----	.046
New Jersey X = .657			----

Note. * p = .05 HSD Critical Value = 0.153

Table 7

Tukey HSD Comparison of Criterion Variable Means

	Course Grades X = .114	GPA X = .156	Alternate Test X = .657
Course Grades X = .114	----	.042	.478*
GPA X = .156		----	.438*
Alternate Tests X = .592			----

Note. * p = .05 HSD Critical Value = 0.055

RESEARCH QUESTIONS	RESULTS (Pearson)	(ANOVA)	Tukey HSD	CONCLUSIONS	DISCUSSION
1. Are the validity coefficients related to the sample size?	r(16) =.029 cv = .4438 at alpha=.05 [NO]				Replicate: --record an accurate N; --use similar size populations
2. Are the validity coefficients related to the type of subtests?		F = .853 cv = 3.18 at alpha = .05 [NO]			Replicate: --give same subtests consistently
3. Are the validity coefficients related to the location of the testing sites?		F = 6.689 cv = 3.63 at alpha = .05 [YES]	National is significantly different from Arizona New Jersey	Obvious differences-- size of N --criterion --Subtests used	Replicate: --Record data by individual state & make comparative analysis --calculate national validity coefficient
4. Are the validity coefficients related to the criterion variable?		F = 113.168 cv = 3.68 at alpha = .05 [YES]	Alternate Tests are significantly different from GPA Course grades	Look again at GPA and grades	Replicate: --Consistently use same criterion variable: • C Form • GPA/grades